



Published in final edited form as:

Addict Biol. 2015 May ; 20(3): 617–627. doi:10.1111/adb.12153.

## Association of substance dependence phenotypes in the COGA sample

Leah Wetherill<sup>1</sup>, Arpana Agrawal<sup>2</sup>, Manav Kapoor<sup>2</sup>, Sarah Bertelsen<sup>2</sup>, Laura J. Bierut<sup>2</sup>, Andrew Brooks<sup>3</sup>, Danielle Dick<sup>4</sup>, Michie Hesselbrock<sup>5</sup>, Victor Hesselbrock<sup>5</sup>, Daniel L. Koller<sup>1</sup>, Nhung Le<sup>2</sup>, John I. Nurnberger Jr.<sup>1</sup>, Jessica E. Salvatore<sup>4</sup>, Marc Schuckit<sup>6</sup>, Jay A. Tischfield<sup>3</sup>, Jen-Chyong Wang<sup>2</sup>, Xiaoling Xuei<sup>1</sup>, Howard J. Edenberg<sup>1</sup>, Bernice Porjesz<sup>7</sup>, Kathleen Bucholz<sup>2</sup>, Alison M. Goate<sup>2</sup>, and Tatiana Foroud<sup>1</sup>

<sup>1</sup>Indiana University School of Medicine

<sup>2</sup>Washington University School of Medicine

<sup>3</sup>Rutgers University

<sup>4</sup>Virginia Commonwealth University

<sup>5</sup>University of Connecticut Health Center

<sup>6</sup>University of California, San Diego

<sup>7</sup>SUNY Health Sciences Center

### Abstract

Alcohol and drug use disorders are individually heritable (50%). Twin studies indicate that alcohol and substance use disorders share common genetic influences, and therefore may represent a more heritable form of addiction and thus be more powerful for genetic studies. This study utilized data from 2,322 subjects from 118 European-American families in the COGA sample to conduct genomewide association analysis of a binary and a continuous index of general substance dependence liability. The binary phenotype (ANYDEP) was based on meeting lifetime criteria for any DSM-IV dependence on alcohol, cannabis, cocaine or opioids. The quantitative trait (QUANTDEP) was constructed from factor analysis based on endorsement across the 7 DSM-IV criteria for each of the 4 substances. Heritability was estimated to be 54% for ANYDEP and 86% for QUANTDEP. One SNP, rs2952621 in the uncharacterized gene *LOC151121* on chromosome 2, was associated with ANYDEP ( $p=1.8 \times 10^{-8}$ ), with support from surrounding imputed SNPs and replication in an independent sample (SAGE;  $p=0.02$ ). One SNP, rs2567261 in *ARHGAP28* (Rho

---

**Address correspondence to:** Tatiana Foroud, Ph.D., Department of Medical and Molecular Genetics, Indiana University School of Medicine, 410 W. 10<sup>th</sup> Street (HS4000), Indianapolis, IN 46202, Tel: 317-278-1291, Fax: 317-278-1100, tforoud@iu.edu.

### AUTHORS CONTRIBUTION

The authors LW, AA, KB, and TF were responsible for the study concept and design, and drafted the manuscript. MH, VH, JIN, MS, LB, and BP were responsible for sample acquisition and characterization. AMG, J-CW, MK, SB, NL were responsible for acquisition of genotype data. LW, AA, and MK conducted statistical analysis. AB, DD, DLK, JES, JAT, XX, HJE, BP provided critical revision of the manuscript for important intellectual content. All authors critically reviewed content and approved final version for publication. Laura J. Bierut, Alison Goate, and Jen-Chyong Wang are listed as an inventors on Issued U.S. Patent 8080,371, "Markers for Addiction" covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction.

All other authors report no biomedical financial interests or potential conflicts of interest.

GTPase activating protein 28), was associated with QUANTDEP ( $p=3.8\times 10^{-8}$ ), and supported by imputed SNPs in the region, but did not replicate in an independent sample (SAGE;  $p=0.29$ ). The results of this study provide evidence that there are common variants that contribute to the risk for a general liability to substance dependence.

### Keywords

alcohol dependence; cannabis dependence; cocaine dependence; common genetic liability; drug dependence; opioid dependence

## INTRODUCTION

An estimated 15.3 million adults in the United States met criteria for an alcohol use disorder in the past 12 months. Of those with alcohol use disorders, 2.3 million adults also met criteria for a drug use disorder (Stinson et al., 2005) with odds ratios estimated to be 7.4 for any drug use disorder, but 3.4 to 19.2 for specific drug use disorders (Khan et al., 2013; Stinson et al., 2005). Both alcohol and drug use disorders are heritable, with approximately 50% of the variance attributable to heritable factors (Bienvenu et al., 2011; Ducci and Goldman, 2012; Tsuang et al., 1998; Wang et al., 2012), although this estimate varies dramatically by substance (e.g. up to 70% heritability for cocaine dependence) (Kendler et al., 2000), age (Bergen et al., 2007; Derringer et al., 2008; Vrieze et al., 2012) and other characteristics, including comorbid psychopathology (Pickens et al., 1991).

This heritable variation can be parsed into those genetic influences that are specific to each drug and importantly, those genetic factors that confer a general predisposition to alcohol and/or substance use disorders, and even other disinhibited behaviors (Hicks et al., 2013; Krueger et al., 2002). Two large twin studies have convincingly shown that a preponderance of the genetic factors influencing illicit drug use disorders overlap (Kendler et al., 2003; Tsuang et al., 2001). Noticeably, when these models were extended to include alcohol use disorders, there was evidence for highly correlated genetic factors ( $r=0.82$ ) that individually influenced the covariation in alcohol and nicotine dependence as well as cannabis and cocaine dependence (Kendler et al., 2007). The extent of genetic overlap was strong for some substances - for instance, 55% and 24% of the genetic variance in alcohol dependence was due to the licit and illicit drug factors respectively, with the remainder being substance-specific. In contrast, for nicotine dependence, 63% of the genetic variance was drug-specific (with 26% and 11% attributable to heritable variation in the licit and illicit factors respectively). Similar to the individual heritability of each substance, there is growing evidence that the heritable covariation across substances changes across development (Vrieze et al., 2012; Young et al., 2006). Irrespective of development and substance-specific variation, there is broad consensus that gene discovery efforts targeting aggregate genetic variation that indexes a shared liability to a variety of substance use disorders, as well as disinhibition, can be profitable (McGue et al., 2013; Vrieze et al., 2013a), with one study showing evidence for genomewide pleiotropic effects across substance use disorders (Vrieze et al., 2013b).

There are multiple approaches, both phenotypic and genetic, to capture the commonality underlying alcohol and substance use disorders and the present study utilizes two straightforward phenotypic approaches. We opted for simple dependence-based phenotypic traits as they lend themselves to replication and future meta-analysis. First, we utilized a binary phenotype, with affection status defined as meeting dependence criteria for at least one substance (alcohol, cannabis, cocaine or opioids), termed *ANYDEP*. Second, we used factor analysis to combine dependence criteria across substances into a continuous quantitative trait representing vulnerability to multiple substance dependence, termed *QUANTDEP*. This quantitative measure is heritable (approx. 60%) (Palmer et al., 2012) and has previously been used in genomic studies (Yang et al., 2012), the most recent of which utilized a similar expanded factorial measure of behavioral disinhibition (including alcohol, nicotine, cannabis and other illicit drug use disorders) to conduct genomewide association and rare nonsynonymous variant analyses (McGue et al., 2013; Vrieze et al., 2013a). These studies did not identify any single common or rare variant at a genomewide significant level; however, the authors reported that 84% of the heritability in illicit drug use was explained by both common and rare variants. While the work of McGue and colleagues included multiple measures of nicotine use and dependence, we elected to exclude nicotine from these measures of general liability based on the work by Kendler and colleagues (Kendler et al., 2007), which showed significant drug-specific genetic influences on nicotine dependence.

In this study, we utilized data from 2,322 subjects from 118 families of European-American descent ascertained for alcohol dependence liability to conduct genomewide association analysis of a binary and a continuous index of general substance dependence liability. While some prior genomewide efforts (McGue et al., 2013) have utilized similar phenotypes in population samples of related individuals, the ascertainment strategy and extended family-based design in our study should increase our ability to detect genetic variation in this phenotype. First, there is substantial evidence that alcohol use disorders that co-aggregate with other substance use disorders (Khan et al., 2013) may represent a more heritable form of addiction (Pickens et al., 1991). Secondly, by modeling the strength of the phenotypic correlation across different degrees of genetic relatedness (i.e. kinship), we utilize data on all related individuals, even those not meeting criteria for diagnoses, allowing us to better explore the extent of co-aggregation of genetic risk across alcohol, cannabis, cocaine and opioid dependence.

## Materials and methods

### Sample

Six sites participating in the Collaborative Study on the Genetics of Alcoholism (COGA) (Begleiter et al., 1995; Foroud et al., 2000) recruited alcohol dependent probands from inpatient and outpatient facilities. The probands and their family members were administered a poly-diagnostic interview, the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) (Bucholz et al., 1994; Hesselbrock et al., 1999). Individuals 17 years of age or younger were administered an adolescent version of the SSAGA. Institutional review boards at all sites approved the study.

A subset of the most genetically informative families was selected for a family-based genome-wide association analysis (GWAS). This sample has been described in detail elsewhere (Kang et al., 2013; Wang et al., 2013) but salient characteristics are presented here. Families were prioritized based on the number of family members with: (1) available DNA who were also alcohol dependent, (2) available DNA who also had electrophysiology data, and (3) available DNA, regardless of other phenotypes. To reduce heterogeneity, only families consisting primarily of self-reported European American ethnicity were included in the sample. The final sample was comprised of 118 large European American families consisting of 2,322 individuals with available DNA.

## Phenotypes and Statistical Analysis

Phenotype data for four substances (alcohol, cannabis, cocaine, and opioids) were obtained from the SSAGA. Some individuals were assessed more than once, in which case data from the SSAGA interview at which an individual reported the maximum number of DSM-IV criteria endorsements for the particular substance was used. Two phenotypes were used in the genetic analyses: ANYDEP, a binary aggregate substance dependence phenotype, and QUANTDEP, a quantitative (continuous) substance dependence phenotype developed using factor analysis.

For ANYDEP, individuals were considered affected if they met DSM-IV lifetime dependence criteria for any of the 4 substances, and unaffected if they did not meet DSM-IV dependence criteria for all 4 drugs. Individuals younger than 23 years old at their most recent interview who did not meet criteria for dependence on any of the 4 drugs were recoded to missing/unknown (n=408) because they had not passed through the primary age of risk. Selection of this age cutoff was based on the median age of onset of alcohol, cannabis, cocaine and opioid/heroin dependence in the White subsample of the U.S. population based National Epidemiologic Survey of Alcohol and Related Conditions (NESARC) (Grant et al., 2004; Hingson et al., 2006). The median ages ranged from 18-22 years, supporting a cutoff of 23 years. In addition, those individuals with insufficient SSAGA to determine whether they were or were not dependent were also coded as unknown (n=144).

QUANTDEP, the quantitative factor score, was constructed by conducting a confirmatory factor analysis of the 7 DSM-IV lifetime dependence criteria (coded as present or absent) for each of the 4 substances (28 items total). As we were interested in those genetic underpinnings that were common to all dependence criteria across the four substances, we elected to use a single factor confirmatory model and did not conduct exploratory analyses, in addition to limiting the factor analysis to the dependence criteria to exclude abuse. All individuals with DSM-IV criteria data were utilized, regardless of age, or substance use. The factor score from the resulting confirmatory analyses was utilized as the quantitative phenotype.

Heritability was estimated for the two phenotypes using the polygenic option in SOLAR (Almasy and Blangero, 1998). The correlation between the total number of DSM-IV criteria endorsed (between 0 and 28) and QUANTDEP was estimated using the Pearson correlation coefficient. Analysis of variance (ANOVA) was used to test if QUANTDEP differed

according to the number of substance dependence diagnoses met. We also tested if the average QUANTDEP value differed across alcohol, cannabis, cocaine and opioid dependence diagnoses. Post-hoc pairwise comparisons employed a Tukey correction for multiple testing.

### Genotyping and Association Analysis

Genotyping for 2,105 subjects in these 118 families was performed at the Genome Technology Access Center at Washington University School of Medicine in St. Louis using the Illumina Human OmniExpress array 12.VI (Illumina, San Diego, CA, USA). In addition, genotypes previously generated on the Illumina Human 1M-Duo BeadChip (Illumina, San Diego, CA, USA) by the Center for Inherited Disease Research were included for 224 subjects from these families (Edenberg et al., 2010). Further details describing data cleaning can be found in Wetherill et al. (Wetherill et al., 2014). The final analytic sample included 2,322 genotyped individuals. This yielded an average of 19.6 genotyped members per family.

The Genome-Wide Association Analysis with Family Data package was utilized to analyze ANYDEP, implemented as a logistic regression model. Relatedness between family members was accounted for via generalized estimating equations. QUANTDEP was analyzed using a linear mixed effects model as implemented in the kinship library (lmekin) in R (<http://www.inside-r.org/packages/cran/kinship/docs/print.lmekin>). This model in the kinship function allows for the covariance matrix to be completely specified for the random effects. The result is that each family has a different covariance pattern based on the kinship coefficients, to model the familial genetic random effects. Gender and birth cohort defined by year of birth (<1930, 1930-1949, 1950-1969, and 1970), were included as covariates in all analyses described above, including statistical models of association, to account for secular trends (Grucza et al., 2008). As needed, genomic control was applied to correct for inflation. To reduce the scope of multiple testing, only genotyped SNPs were included in the initial analyses. After correcting for the final number of autosomal SNPs ( $n=591,785$ ), the genome-wide significance threshold was  $p=8.45 \times 10^{-8}$ .

In regions with significant association results, we analyzed imputed SNPs to further evaluate the evidence for association. SNPs were imputed to 1000 Genomes (EUR, August 2010 release) using BEAGLE 3.3.1 (Browning and Browning, 2009) as described in Wang et al (Wang et al., 2013). Secondary analyses were performed for significant SNPs to test whether the observed genetic association could be attributed to dependence on a specific substance. Analyses for ANYDEP were performed using the Mantel-Haenszel Chi-Square test of association assuming an additive genetic model. Analyses for QUANTDEP were performed using analysis of covariance employing substance, genotype, and the substance\*genotype interaction to test for differences in genotype by substance. Gender and birth cohort were included as covariates.

### Replication Study

Independent replication of SNPs demonstrating evidence of significant association in the COGA sample was evaluated in the Study of Addiction: Genetics and Environment (SAGE)

sample. SAGE is a case-control sample comprised of three complementary studies: COGA, the Family Study of Cocaine Dependence and the Collaborative Genetics Study of Nicotine Dependence (Bierut, 2010). There were 129 individuals from the 118 COGA families in the current study that overlapped with the SAGE sample, and were removed from the SAGE replication dataset. The remaining independent SAGE sample used for replication was limited to 2,647 individuals of European American descent. Factor analysis scores from Mplus were independently estimated for this study, as described above, based on DSM-IV dependence criteria for the 4 substances. Analyses were implemented in Plink and included age at interview and gender as covariates.

## RESULTS

### ANYDEP

The number of individuals utilized for the categorical phenotype ANYDEP was 1,770 (59% male). Summary results are provided only for genotyped individuals for both phenotypes. Nearly half the sample met DSM-IV criteria for at least one substance (ANYDEP,  $N=832$ , 47%). The COGA sample was ascertained through an alcohol dependent individual in treatment and families were selected for the highest density of alcohol dependent members; therefore, it was expected that there would be many individuals meeting criteria for alcohol dependence (40% in the full sample, but 84% of those who were dependent on alcohol and/or any other substance). In addition, 19% met criteria for cannabis dependence (with or without alcohol dependence). The rates for cocaine and opioid dependence were lower (11% and 4% respectively). There were 832 individuals that met criteria for at least one substance dependence diagnosis; of those, 312 (37.5%) endorsed at least two diagnoses. Alcohol and cannabis dependence were the most common (15%; Table 1).

### QUANTDEP

The number of individuals included in the analysis of the quantitative factor score QUANTDEP was 2,183 (47% male). The confirmatory one factor model fit the data well in COGA (CFI=0.96; RMSEA=0.07) and in the replication sample, SAGE (CFI=0.98; RMSEA=0.07), supporting our proposed unidimensional conceptualization of dependence criteria for alcohol, cannabis, cocaine and opioids. Factor loadings in the COGA sample ranged from 0.67 (alcohol dependence criterion *desire to cut down*) to 0.99 (for several cocaine dependence criteria) and were highly consistent across COGA and SAGE. In general, factor loadings for alcohol and cannabis dependence criteria were lower, and ranged between 0.67-0.85 while those for cocaine and opioids were uniformly high (0.84-0.99). Data across drug classes and across criteria loadings (and standard errors) for all 7 criteria for the 4 substances are available in Supplemental Material Figure 1.

### ANYDEP and QUANTDEP

As expected, individuals with no substance dependence had the lowest QUANTDEP values (mean =  $-0.40$ , standard error (se) =  $0.01$ ) and those with a dependence diagnosis for all 4 substances had the highest QUANTDEP values (mean =  $1.45$ , se =  $0.06$ ; overall  $F(8,1762)=659.4$ ,  $p<0.0001$ ; all pairwise Tukey-adjusted  $p<0.006$ ) (Figure 1A). QUANTDEP values differed based on the substance (overall  $F(7,502)=32.9$ ,  $p<0.0001$ ; all



pairwise Tukey-adjusted  $p < 0.03$ ), with individuals meeting cocaine dependence criteria having the highest mean QUANTDEP scores and those meeting criteria for alcohol dependence having the lowest (Figure 1B) of individuals meeting criteria for dependence on any substance. QUANTDEP was positively correlated with the total number of DSM-IV criteria endorsed across the 4 substances ( $\rho = 0.88$ ,  $p < 0.0001$ ).

**Genetic Analysis**—The ANYDEP phenotype was moderately heritable, with an estimate of 0.54 (standard error = 0.08,  $p = 1.3 \times 10^{-17}$ ). Heritability for the QUANTDEP was higher, with an estimate of 0.86 (standard error = 0.02,  $p = 2.1 \times 10^{-66}$ ).

Association analyses results are summarized in Figure 2. All association results with  $p < 1.0 \times 10^{-6}$  are shown in Table 2. More extensive results are provided in Supplemental Table 1. The QQ plot for ANYDEP revealed inflation of p-values ( $\lambda = 1.0608$ ); therefore Genomic Control (GC) p-values were calculated. The lambda for the inflation-corrected p-values = 0.99999, indicating no further inflation. All p-values reported for ANYDEP are GC p-values. There was no evidence to suggest inflation of the association p-values for QUANTDEP (QUANTDEP  $\lambda = 1.0095$ ). QQ plots are provided for each phenotype (Supplemental Figure 2).

Two genotyped SNPs reached genome-wide significance ( $p < 8.4 \times 10^{-8}$ ). A SNP in the uncharacterized gene *LOC151121* on chromosome 2, rs2952621, was associated with ANYDEP ( $p = 1.8 \times 10^{-8}$ ; OR = 1.07). Secondary analysis was performed to test whether this SNP demonstrated greater evidence of association with a particular substance. Although more individuals were alcohol dependent, the pattern of dependence by genotype was similar for all 4 substances ( $\chi^2_{(1)} = 0.48$ ,  $p = 0.49$ ; Figure 3A), indicating that the association was not driven by dependence on one particular substance. Individuals with one or two copies of the minor allele (T) were more likely to be dependent on at least one substance than those having no copies of the minor allele. Analysis of imputed SNPs in this region provided additional evidence to support the association (Figure 4A). There was modest replication for this SNP in the SAGE sample ( $p = 0.02$ , OR = 1.1), with T as the risk allele in both samples. Corresponding SAGE results for the SNPs are provided in Table 2.

The second genome-wide significant finding was observed with QUANTDEP and a genotyped SNP rs2567261 on chromosome 18 in the gene *ARHGAP28* (Rho GTPase activating protein 28) ( $p = 3.8 \times 10^{-8}$ ). Due to the low minor allele frequency for this SNP (MAF=0.08), AA and AG genotypes were combined. Although individuals dependent on opioids and cocaine had higher QUANTDEP scores on average (main effect of substance  $p < 0.0001$ ; Figure 3B), there was no substance\*genotype effect ( $p = 0.74$ ) confirming that QUANTDEP exhibited the same pattern by genotype across the 4 substances. Analysis of the imputed SNPs in this region further supported the association (Figure 4B). There was no evidence of replication in SAGE for QUANTDEP ( $p = 0.29$ ). Corresponding SAGE results for the SNPs are provided in Table 2.

## DISCUSSION

This is one of the first GWAS to test for the association of overall substance dependence phenotypes, defined both categorically (dependence diagnosis for alcohol, cannabis, cocaine or opioids; ANYDEP) and quantitatively (factor analysis of the 7 DSM-IV dependence criteria, across alcohol, cannabis, cocaine and opioids; QUANTDEP). This approach implicitly tested the hypothesis that there are genes with pleiotropic effects contributing to dependence on alcohol, cannabis, cocaine and opioids. Using these multi-substance phenotypes, we detected genome-wide significant results with SNPs in two different genes. This finding is consistent with an extensive twin literature that provides demonstrable support for common genetic liability underlying addiction to multiple substances (Kendler et al., 2003; Tsuang et al., 2001). Furthermore, a previous study in a slightly different COGA sample demonstrated aggregation of drug dependence in relatives of alcohol-dependent probands, even after controlling for comorbidity in the probands (Nurnberger et al., 2004).

Genomewide significant association for ANYDEP was observed with a SNP in an uncharacterized gene, *LOC151121* ( $p=1.8\times10^{-8}$ ). Further evidence of association was corroborated by surrounding SNPs, both genotyped and imputed. Nominal replication was found in the SAGE sample ( $p=0.02$ ) with the same phenotype. This SNP was moderately associated with QUANTDEP ( $p=3.8\times10^{-5}$ ) and also with the number of DSM-IV alcohol dependence criteria endorsed (symptom count) in another related study with data from the same sample ( $p=7.2\times10^{-5}$ ), (Wang et al., 2013). Similar to the replication results here, this SNP was nominally associated with the alcohol dependence symptom count in the SAGE sample as well ( $p=0.014$ ) (Wang et al., 2013).

Significant association was also detected with QUANTDEP for the SNP rs2567261 in *ARHGAP28* ( $p=3.8\times10^{-8}$ ). Further evidence of association was observed with both genotyped and imputed SNPs within the gene. *ARHGAP28* is also known as Rho GTPase activating protein 28. GTPase-activating proteins target GTPases, and are mediated by exposure to alcohol, cannabis, cocaine, and opioids. For example Rho1 and Rac moderate the stimulating and sedative effects of acute ethanol intoxication in *Drosophila* (Rothenfluh et al., 2006). Thus, there is strong biological rationale for this gene as a potential candidate for substance dependence. Of note, this SNP was modestly associated with ANYDEP ( $p=3.2\times10^{-4}$ ) in this sample and with previously published alcohol symptom count in the COGA family sample ( $p=8.7\times10^{-5}$ ). However, rs2567261 was not significantly associated with alcohol symptom count in the SAGE sample, although there was a trend in that direction ( $p=0.08$ ) (Wang et al., 2013). Although the association did not replicate in the SAGE sample for this phenotype, there was a trend towards association with the other phenotype, ANYDEP ( $p=0.08$ ). This weak replication for a different phenotype may be due to the fact that the majority of the SAGE sample was ascertained on nicotine and cocaine dependence, whereas the COGA sample was recruited based on an alcohol dependent proband and expanded to include the maximum number of alcohol dependent family members. The exclusion of nicotine dependence criteria may have attenuated the likelihood of replication, given ascertainment for nicotine dependence in SAGE. Since the SNP association was not primarily due to alcohol dependence, it is possible that high rates of comorbidity in the COGA sample as compared with the SAGE sample contributed to the



finding. In addition, the family-based test in the large COGA families with comorbidity may have had greater power to detect the association. QUANTDEP seems to represent an underlying severity of addiction. As seen in Figure 3A, the higher the number of comorbid diagnoses, the higher the QUANTDEP scores.

Overall, for the QUANTDEP measure, the loadings for the alcohol and cannabis criteria appear generalizable to general population studies (Lynskey and Agrawal, 2007; Saha et al., 2006); in contrast, those for cocaine and opioid dependence are higher (typically  $> 0.9$ ). In COGA, cocaine (7-12%) and opioid (2-5%) dependence criteria were less commonly endorsed than those for alcohol (15-51%) and cannabis (9-19%), with the former also showing less range in endorsement rates (i.e. each dependence criterion was equally likely to be endorsed). In SAGE, cocaine dependence criteria (17-19%) were somewhat more commonly endorsed than cannabis (12-18%) dependence criteria, yet the cocaine criteria had higher loadings than the cannabis criteria, identical to COGA. For both cocaine and opioids, the range of prevalence of individual criteria was highly restricted (e.g. 6-7% for opioid criteria). Thus, in both COGA and SAGE, the likelihood of endorsement of each of the 7 dependence criteria for cocaine and opioids was similar while certain alcohol and cannabis criteria (e.g. *tolerance*) were endorsed more often than others (e.g. *use dominates life*). This may be related to the ascertainment strategy and overrepresentation of family history for alcoholism in both samples. Nonetheless, all factor loadings were high, indicating that QUANTDEP reflects a general liability to dependence across multiple substances. In particular, QUANTDEP captures the liability to cocaine and opioid dependence criteria in these two studies. Therefore, in addition to being an index of severity and a measure of general liability to addiction across alcohol and drugs, QUANTDEP likely also reflects variation in prevalence and the expected pattern of comorbid relationships and co-aggregation across alcohol and drug dependence criteria in these subjects ascertained on specific substance dependence.

While the heritability of the binary phenotype of dependence on any substance was similar in this sample ( $h^2=0.54$ ) to the most common heritability estimate ( $h^2=0.50$ ) reported for any of the four substances from twin studies (Goldman et al., 2005), the heritability for the quantitative phenotype was much higher ( $h^2=0.86$ ). This is consistent with one prior twin study of a latent genetic factor ( $h^2=0.81$ ) underlying alcohol and drug problems as well as measures of impulsivity and conduct problems (Krueger et al., 2002) but significantly higher than some others (e.g.  $h^2=0.40$ ) (Button et al., 2009). Although this high heritability should not be over-interpreted (Goldman et al., 2005), it is possible that the use of a multi-variable quantitative phenotype, utilizing the pattern of endorsement of the 7 DSM-IV criteria across all four substances, captured valuable genetic information across the vulnerability spectrum.

The two phenotypes used in this study were both aggregate measures of overall dependence. Although their top genetic signals did not overlap (i.e. the same SNP did not reach statistical significance for both phenotypes), there was evidence of association for the other phenotype for the two SNPs that attained genomewide significance (see Supplemental Table 1 for additional comparisons). The difference in magnitude of p-value is not surprising given the arguable validity of the diagnostic cutoff (i.e. 3 or more of 7 dependence criteria) implemented in DSM-IV which likely excluded from affected status (for ANYDEP), a

number of individuals who may have met criteria for abuse or endorsed 1-2 dependence symptoms across one or even multiple drugs, and thus did not qualify for dependence. Viewed alternatively, the unaffected individuals for ANYDEP represent a heterogeneous group varying in severity. Such variability was better captured by QUANTDEP, which while not taking abuse criteria into account, was a better approximation of the range of vulnerability to substance-related problems. Thus, it is likely that ANYDEP reflects the more severe of the QUANTDEP scores. Finally, the possibility that our findings reflect false positives cannot be excluded.

There have been multiple prior GWAS that have utilized symptom counts and factor scores of alcohol dependence criteria (Vrieze et al., 2013a; Yang et al., 2012) but only one attempted to combine indices of alcohol (consumption and dependence), nicotine and drug misuse (with disinhibition measures) using hierarchical factorial analyses for GWAS. In that study, McGue and colleagues (McGue et al., 2013) reported on 4 SNPs associated with multiple first order (e.g. alcohol consumption, alcohol dependence, illicit drug dependence) and higher order externalizing factors. One of these SNPs, rs10037670 in *GALNT10*, with the highest association for illicit drug dependence factor ( $p=3.8\times10^{-6}$ ) was modestly associated in this study with both ANYDEP ( $p=0.0029$ ) and QUANTDEP ( $p=0.0067$ ).

There are several strengths of this study design, the first being the use of families densely affected with alcohol dependent individuals. Family (Merikangas et al., 1998; Nurnberger et al., 2004) and twin (Tsuang et al., 1998) studies suggest familial co-aggregation and heritable overlap across alcohol, cannabis, cocaine and opioids. Thus, this family-based COGA sample, enriched for dependence on multiple substances and shared genetic risk, allowed us to test for the association of common variants with risk for dependence across multiple substances. A second strength of this study was the use of a family-based association design. This allowed us to examine association within a family consisting of members who endorsed criteria for dependence on different (or multiple) substances. Third, family-based analysis is robust to population substructures such as nuanced differences in ethnicity, which might occur with marry-in individuals of a different race, and in turn affects the genetic diversity of the offspring.

Three caveats are worth considering. First, only a small subset of individuals met DSM-IV criteria for opioid and cocaine dependence. Thus, it is possible that results pertain more closely to lower liabilities to these substances. Second, we did not include nicotine dependence criteria in this analysis. As we were interested in a confirmatory model of unidimensional genetic risk, we elected to exclude nicotine symptoms based on published evidence for a preponderance of non-overlapping genetic influences on these criteria. Finally, we elected not to utilize abuse criteria (nor craving), despite DSM-5 related changes. Previous findings in COGA families demonstrated that abuse did not aggregate in relatives of alcohol dependent probands (Nurnberger et al., 2004). In addition, the extant psychometric literature suggests that with the exception of hazardous use which is frequently endorsed to the exclusion of other abuse or dependence criteria, the remaining abuse criteria (failure to fulfill role obligations and social/interpersonal problems) and craving are infrequently endorsed in the absence of co-occurring dependence criteria. This is particularly true in samples ascertained for substance use disorders, such as SAGE and

COGA. For instance, in SAGE, of those who reported no alcohol dependence criteria, only 12-26 individuals endorsed at least one abuse criterion other than hazardous use (which was endorsed by 140). Hence, it is unlikely that the exclusion of abuse criteria resulted in our inability to capture a relevant portion of the liability continuum.

In summary, this study provides evidence that there are common variants that contribute to the risk for a general liability to substance dependence, defined qualitatively and quantitatively. The results of this study require replication in independent samples to further explore whether overall dependence on multiple or individual substances is associated with the SNPs in these regions.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

The Collaborative Study on the Genetics of Alcoholism (COGA), Principal Investigators B. Porjesz, V. Hesselbrock, H. J. Edenberg, L. J. Bierut, includes 10 different centers: University of Connecticut (V. Hesselbrock); Indiana University (H. J. Edenberg, J. Nurnberger Jr, T. Foroud); University of Iowa (S. Kuperman, J. Kramer); SUNY Downstate (B. Porjesz); Washington University in St. Louis (L. J. Bierut, A. Goate, J. Rice, K. Bucholz); University of California at San Diego (M. Schuckit); Rutgers University (J. Tischfield); Southwest Foundation (L. Almasy), Howard University (R. Taylor) and Virginia Commonwealth University (D. Dick). Other COGA collaborators include: L. Bauer (University of Connecticut); D. Koller, S. O'Connor, L. Wetherill, X. Xuei (Indiana University); G. Chan (University of Iowa); N. Manz, (SUNY Downstate); J. Rohrbaugh, J.-C. Wang (Washington University in St. Louis); A. Brooks (Rutgers University); and F. Aliev (Virginia Commonwealth University). A. Parsian and M. Reilly are the NIAAA Staff Collaborators.

We continue to be inspired by our memories of Henri Begleiter and Theodore Reich, founding PI and Co-PI of COGA, and also owe a debt of gratitude to other past organizers of COGA, including Ting-Kai Li, currently a consultant with COGA, P. Michael Conneally, Raymond Crowe, and Wendy Reich, for their critical contributions.

This national Collaborative Study is supported by NIH Grant U10AA008401 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse (NIDA). Funding support for GWAS genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the National Institute on Alcohol Abuse and Alcoholism, the NIH GEI (U01HG004438), and the NIH contract "High throughput genotyping for studying the genetic contributions to human disease" (HHSN268200782096C).

AA is supported by K02DA32573 and AAR21235 and JES by F32AA22269.

Funding support for the Study of Addiction: Genetics and Environment (SAGE) was provided through the NIH Genes, Environment and Health Initiative [GEI] (U01 HG004422). SAGE is one of the genome-wide association studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of datasets and samples was provided by the Collaborative Study on the Genetics of Alcoholism (COGA; U10 AA008401), the Collaborative Genetic Study of Nicotine Dependence (COGEND; P01 CA089392), and the Family Study of Cocaine Dependence (FSCD; R01 DA013423, R01 DA019963). Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438), the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse, and the NIH contract "High throughput genotyping for studying the genetic contributions to human disease" (HHSN268200782096C).

## REFERENCES

Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet.* 1998; 62:1198–1211. [PubMed: 9545414]

- Begleiter H, Reich T, Hesselbrock V, Porjesz B, Li TK, Schuckit M, Edenberg H, Rice JP. The Collaborative Study on the Genetics of Alcoholism. *Alcohol Health & Research World*. 1995; 19:228–236.
- Bergen SE, Gardner CO, Kendler KS. Age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood: a meta-analysis. *Twin Res Hum Genet*. 2007; 10:423–433. [PubMed: 17564500]
- Bienvenu OJ, Davydow DS, Kendler KS. Psychiatric ‘diseases’ versus behavioral disorders and degree of genetic influence. *Psychol Med*. 2011; 41:33–40. [PubMed: 20459884]
- Bierut LJ. Convergence of genetic findings for nicotine dependence and smoking related diseases with chromosome 15q24-25. *Trends Pharmacol Sci*. 2010; 31:46–51. [PubMed: 19896728]
- Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet*. 2009; 84:210–223. [PubMed: 19200528]
- Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI Jr. Reich T, Schmidt I, Schuckit MA. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *J Stud Alcohol*. 1994; 55:149–158.
- Button TM, Stallings MC, Rhee SH, Corley RP, Boardman JD, Hewitt JK. Perceived peer delinquency and the genetic predisposition for substance dependence vulnerability. *Drug Alcohol Depend*. 2009; 100:1–8. [PubMed: 19008053]
- Derringer J, Krueger RF, McGue M, Iacono WG. Genetic and environmental contributions to the diversity of substances used in adolescent twins: a longitudinal study of age and sex effects. *Addiction*. 2008; 103:1744–1751. [PubMed: 18821880]
- Ducci F, Goldman D. The genetic basis of addictive disorders. *Psychiatr Clin North Am*. 2012; 35:495–519. [PubMed: 22640768]
- Edenberg HJ, Koller DL, Xuei X, Wetherill L, McClintick JN, Almasy L, Bierut LJ, Bucholz KK, Goate A, Aliev F, Dick D, Hesselbrock V, Hinrichs A, Kramer J, Kuperman S, Nurnberger JI Jr. Rice JP, Schuckit MA, Taylor R, Todd Webb B, Tischfield JA, Porjesz B, Foroud T. Genome-wide association study of alcohol dependence implicates a region on chromosome 11. *Alcohol Clin Exp Res*. 2010; 34:840–852. [PubMed: 20201924]
- Foroud T, Edenberg HJ, Goate A, Rice J, Flury L, Koller DL, Bierut LJ, Conneally PM, Nurnberger JI, Bucholz KK, Li TK, Hesselbrock V, Crowe R, Schuckit M, Porjesz B, Begleiter H, Reich T. Alcoholism susceptibility loci: confirmation studies in a replicate sample and further mapping. *Alcohol Clin Exp Res*. 2000; 24:933–945.
- Goldman D, Oroszi G, Ducci F. The genetics of addictions: uncovering the genes. *Nat Rev Genet*. 2005; 6:521–532. [PubMed: 15995696]
- Grant BF, Dawson DA, Stinson FS, Chou SP, Dufour MC, Pickering RP. The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence: United States, 1991-1992 and 2001-2002. *Drug Alcohol Depend*. 2004; 74:223–234. [PubMed: 15194200]
- Gruza RA, Bucholz KK, Rice JP, Bierut LJ. Secular trends in the lifetime prevalence of alcohol dependence in the United States: a re-evaluation. *Alcohol Clin Exp Res*. 2008; 32:763–770. [PubMed: 18336633]
- Hesselbrock M, Easton C, Bucholz KK, Schuckit M, Hesselbrock V. A validity study of the SSAGA--a comparison with the SCAN. *Addiction*. 1999; 94:1361–1370. [PubMed: 10615721]
- Hicks BM, Foster KT, Iacono WG, McGue M. Genetic and environmental influences on the familial transmission of externalizing disorders in adoptive and twin offspring. *JAMA Psychiatry*. 2013; 70:1076–1083. [PubMed: 23965950]
- Hingson RW, Heeren T, Winter MR. Age at drinking onset and alcohol dependence: age at onset, duration, and severity. *Arch Pediatr Adolesc Med*. 2006; 160:739–746. [PubMed: 16818840]
- Kang, S.; Rangaswamy, M.; Chorlian, D.; Manz, N.; Porjesz, B.; COGA. XXIst World Congress of Psychiatric Genetics. Boston: 2013. Family-based Genome-Wide Association Study of resting EEG identifies UROC1 and neurotransmitter biosynthetic process.
- Kendler KS, Jacobson KC, Prescott CA, Neale MC. Specificity of genetic and environmental risk factors for use and abuse/dependence of cannabis, cocaine, hallucinogens, sedatives, stimulants, and opiates in male twins. *Am J Psychiatry*. 2003; 160:687–695. [PubMed: 12668357]

- Kendler KS, Karkowski LM, Neale MC, Prescott CA. Illicit psychoactive substance use, heavy use, abuse, and dependence in a US population-based sample of male twins. *Arch Gen Psychiatry*. 2000; 57:261–269. [PubMed: 10711912]
- Kendler KS, Myers J, Prescott CA. Specificity of genetic and environmental risk factors for symptoms of cannabis, cocaine, alcohol, caffeine, and nicotine dependence. *Arch Gen Psychiatry*. 2007; 64:1313–1320. [PubMed: 17984400]
- Khan S, Okuda M, Hasin DS, Secades-Villa R, Keyes K, Lin KH, Grant B, Blanco C. Gender differences in lifetime alcohol dependence: results from the national epidemiologic survey on alcohol and related conditions. *Alcohol Clin Exp Res*. 2013; 37:1696–1705. [PubMed: 23763329]
- Krueger RF, Hicks BM, Patrick CJ, Carlson SR, Iacono WG, McGue M. Etiologic connections among substance dependence, antisocial behavior, and personality: modeling the externalizing spectrum. *J Abnorm Psychol*. 2002; 111:411–424. [PubMed: 12150417]
- Lynskey MT, Agrawal A. Psychometric properties of DSM assessments of illicit drug abuse and dependence: results from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *Psychol Med*. 2007; 37:1345–1355. [PubMed: 17407621]
- McGue M, Zhang Y, Miller MB, Basu S, Vrieze S, Hicks B, Malone S, Oetting WS, Iacono WG. A genome-wide association study of behavioral disinhibition. *Behav Genet*. 2013; 43:363–373. [PubMed: 23942779]
- Merikangas KR, Stolar M, Stevens DE, Goulet J, Preisig MA, Fenton B, Zhang H, O'Malley SS, Rounsaville BJ. Familial transmission of substance use disorders. *Arch Gen Psychiatry*. 1998; 55:973–979. [PubMed: 9819065]
- Numberger JI Jr. Wiegand R, Bucholz K, O'Connor S, Meyer ET, Reich T, Rice J, Schuckit M, King L, Petti T, Bierut L, Hinrichs AL, Kuperman S, Hesselbrock V, Porjesz B. A family study of alcohol dependence: coaggregation of multiple disorders in relatives of alcohol-dependent probands. *ArchGenPsychiatry*. 2004; 61:1246–1256.
- Palmer RH, Button TM, Rhee SH, Corley RP, Young SE, Stallings MC, Hopfer CJ, Hewitt JK. Genetic etiology of the common liability to drug dependence: evidence of common and specific mechanisms for DSM-IV dependence symptoms. *Drug Alcohol Depend*. 2012; 123(Suppl 1):S24–32. [PubMed: 22243758]
- Pickens RW, Svikis DS, McGue M, Lykken DT, Heston LL, Clayton PJ. Heterogeneity in the inheritance of alcoholism. A study of male and female twins. *ArchGenPsychiatry*. 1991; 48:19–28.
- Rothenfluh A, Threlkeld RJ, Bainton RJ, Tsai LT, Lasek AW, Heberlein U. Distinct behavioral responses to ethanol are regulated by alternate RhoGAP18B isoforms. *Cell*. 2006; 127:199–211. [PubMed: 17018286]
- Saha TD, Chou SP, Grant BF. Toward an alcohol use disorder continuum using item response theory: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Psychol Med*. 2006; 36:931–941. [PubMed: 16563205]
- Stinson FS, Grant BF, Dawson DA, Ruan WJ, Huang B, Saha T. Comorbidity between DSM-IV alcohol and specific drug use disorders in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Drug Alcohol Depend*. 2005; 80:105–116. [PubMed: 16157233]
- Tsuang MT, Bar JL, Harley RM, Lyons MJ. The Harvard Twin Study of Substance Abuse: what we have learned. *Harv Rev Psychiatry*. 2001; 9:267–279. [PubMed: 11600486]
- Tsuang MT, Lyons MJ, Meyer JM, Doyle T, Eisen SA, Goldberg J, True W, Lin N, Toomey R, Eaves L. Co-occurrence of abuse of different drugs in men: the role of drug-specific and shared vulnerabilities. *Arch Gen Psychiatry*. 1998; 55:967–972. [PubMed: 9819064]
- Vrieze SI, Feng S, Miller MB, Hicks BM, Pankratz N, Abecasis GR, Iacono WG, McGue M. Rare Nonsynonymous Exonic Variants in Addiction and Behavioral Disinhibition. *Biol Psychiatry*. 2013a
- Vrieze SI, Hicks BM, Iacono WG, McGue M. Decline in genetic influence on the co-occurrence of alcohol, marijuana, and nicotine dependence symptoms from age 14 to 29. *Am J Psychiatry*. 2012; 169:1073–1081. [PubMed: 22983309]
- Vrieze SI, McGue M, Miller MB, Hicks BM, Iacono WG. Three mutually informative ways to understand the genetic relationships among behavioral disinhibition, alcohol use, drug use,

nicotine use/dependence, and their co-occurrence: twin biometry, GCTA, and genome-wide scoring. *Behav Genet.* 2013b; 43:97–107. [PubMed: 23362009]

Wang JC, Foroud T, Hinrichs AL, Le NXH, Bertelsen S, Budde J, Chou YL, Harari O, Koller DL, Wetherill L, Agrawal A, Almasy L, Brooks A, Bucholz K, Dick D, Hesselbrock V, Kang S, Kapoor M, Kramer J, Kuperman S, Manz N, McClintick JN, Nurnberger JJ, Ragaswamy M, Rice J, Schuckit M, Tischfield JA, Xuei X, Porjesz B, Heath AC, Edenberg HJ, Bierut LJ, Goate AM. A genome-wide association study of alcohol-dependence symptom counts in extended pedigrees identifies C15orf53. *Mol Psychiatry.* 2013; 18:1218–1224. [PubMed: 23089632]

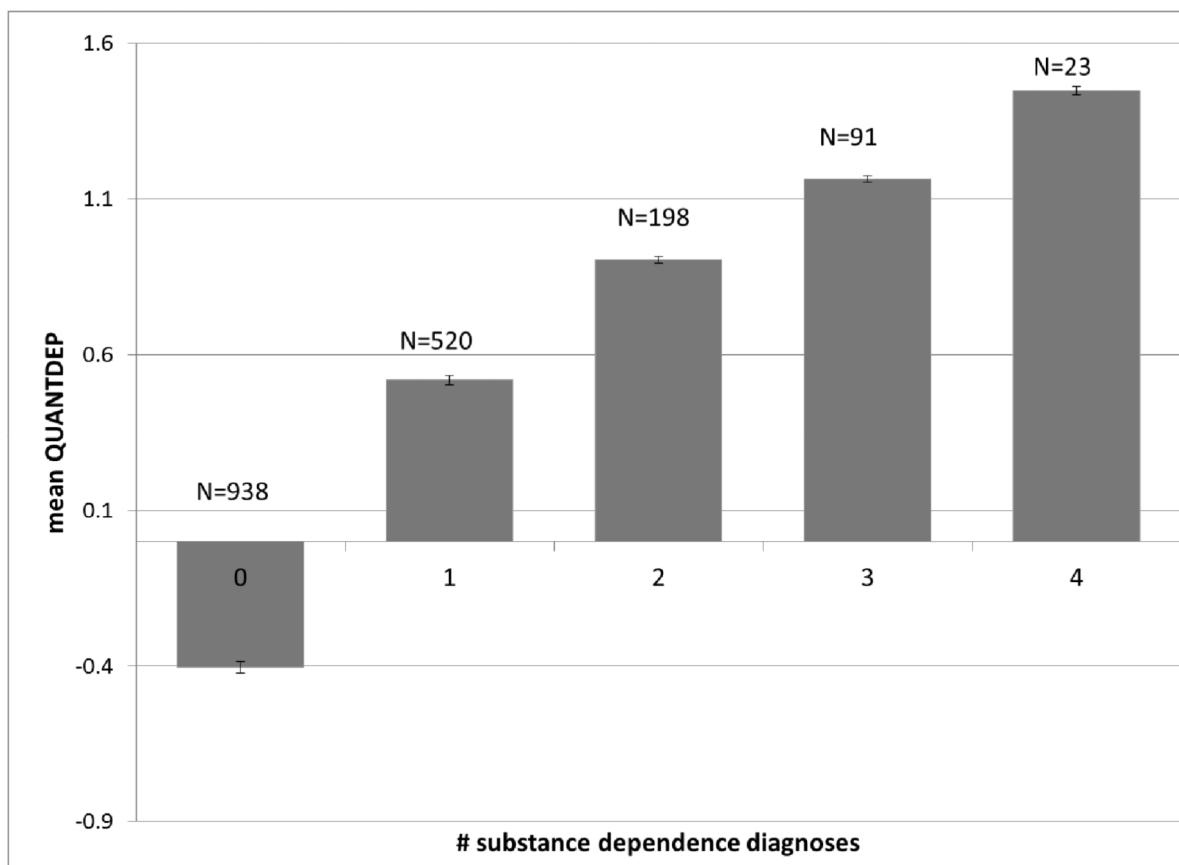
Wang JC, Kapoor M, Goate AM. The genetics of substance dependence. *Annu Rev Genomics Hum Genet.* 2012; 13:241–261. [PubMed: 22703173]

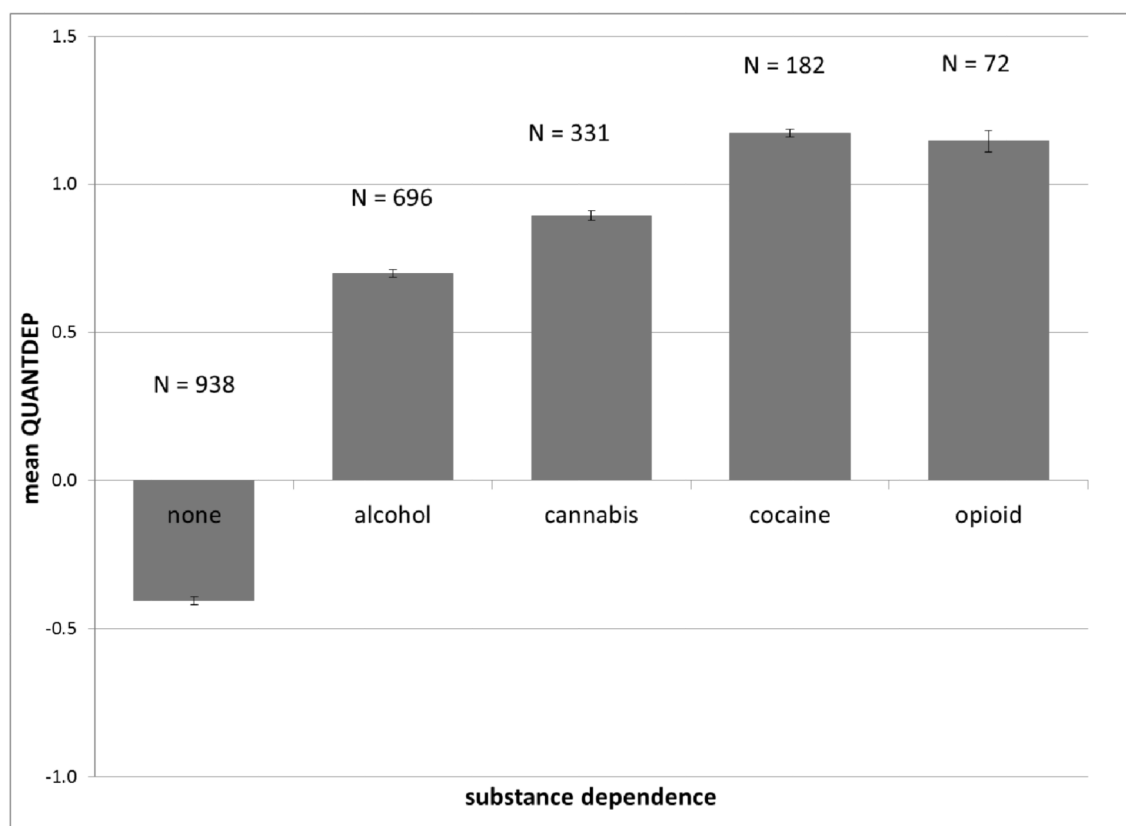
Wetherill L, Kapoor M, Agrawal A, Bucholz K, Koller DL, Bertelsen S, Le NX, Wang JC, Almasy L, Hesselbrock M, Kramer J, Nurnberger J, Schuckit M, Tischfield J, Xuei X, Porjesz B, Edenberg H, Goate A, Foroud T. Family-based association analysis of alcohol dependence criteria and severity. *Alcoholism: Clinical and Experimental Research.* 2014; 38:354–366.

Yang BZ, Han S, Kranzler HR, Farrer LA, Elston RC, Gelernter J. Autosomal linkage scan for loci predisposing to comorbid dependence on multiple substances. *Am J Med Genet B Neuropsychiatr Genet.* 2012; 159B:361–369. [PubMed: 22354695]

Young SE, Rhee SH, Stallings MC, Corley RP, Hewitt JK. Genetic and environmental vulnerabilities underlying adolescent substance use and problem use: general or specific? *Behav Genet.* 2006; 36:603–615. [PubMed: 16619135]

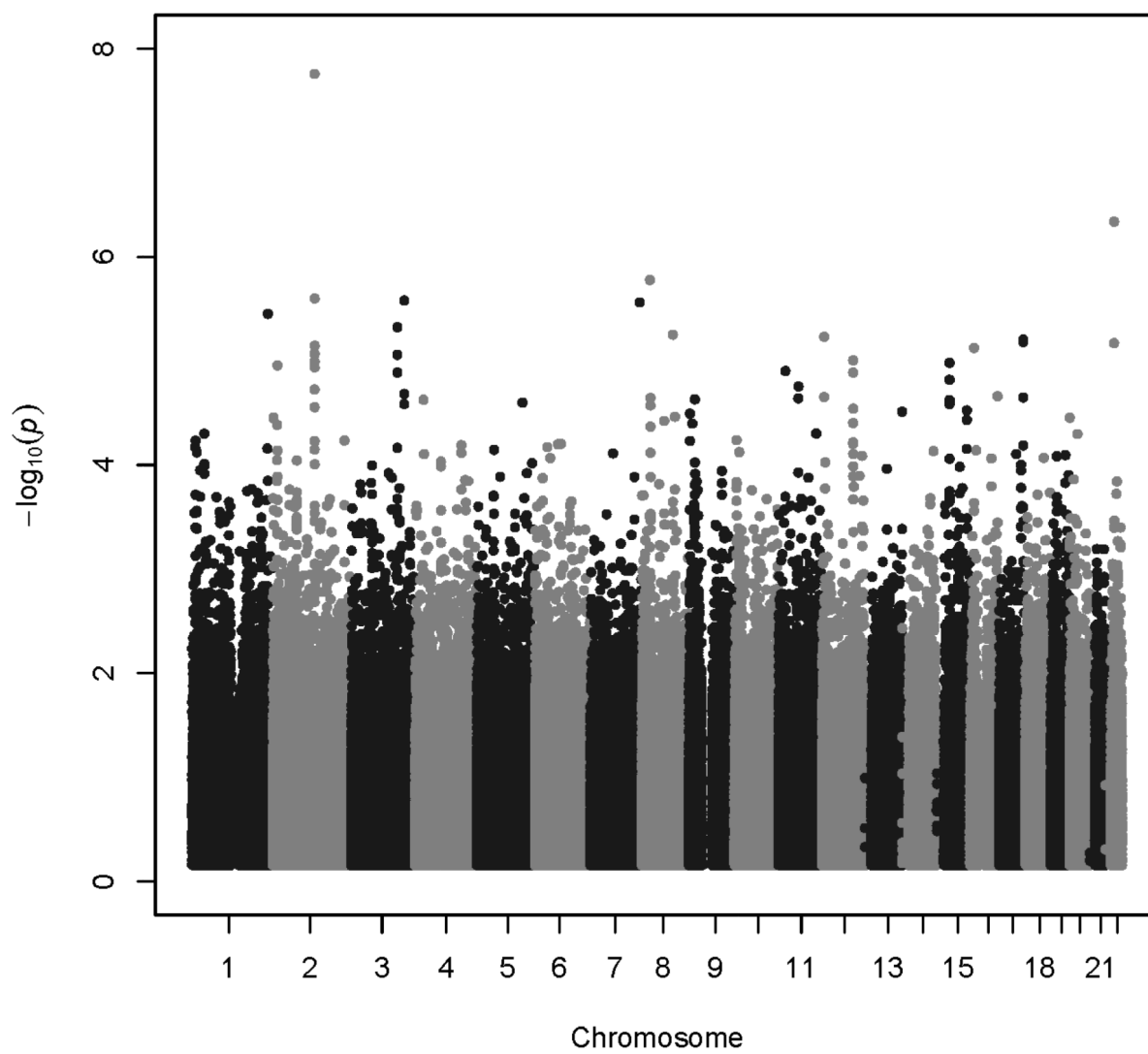


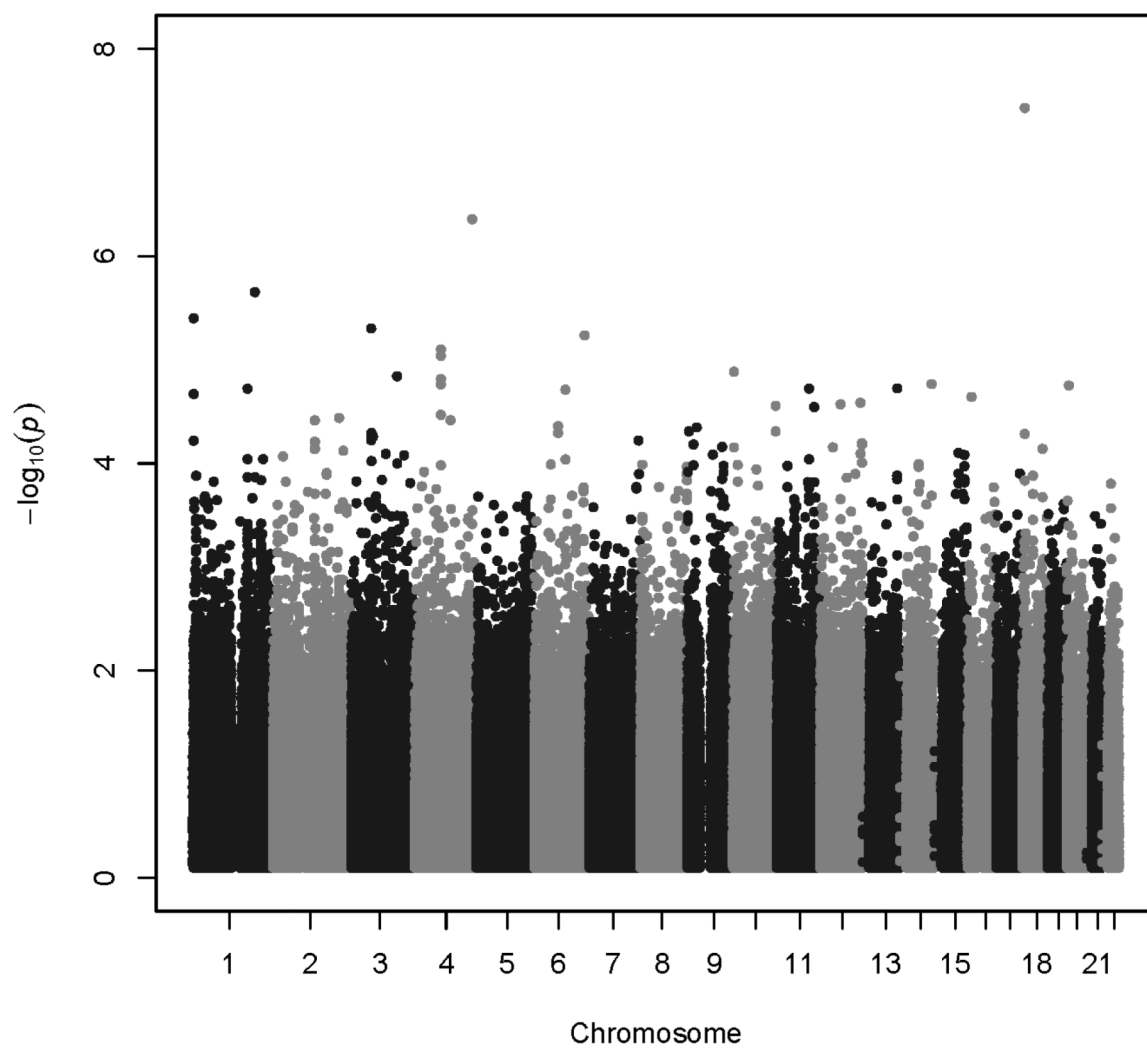




**Figure 1.**

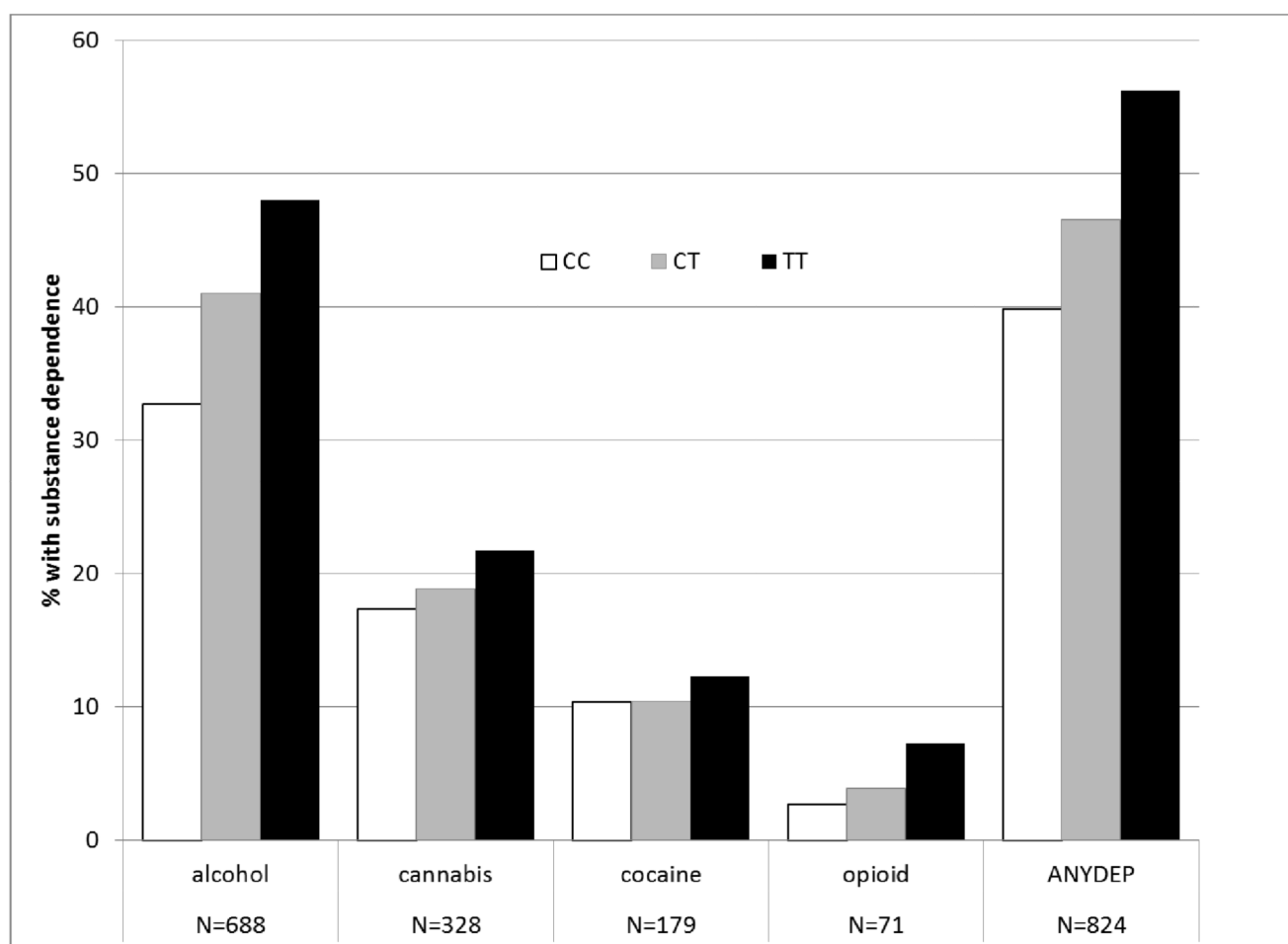
Mean and standard error of ANYQUANT scores for individuals used in analysis based on:  
Number of substance dependence diagnoses; B. Each DSM-IV substance dependence  
endorsed (includes comorbidities)

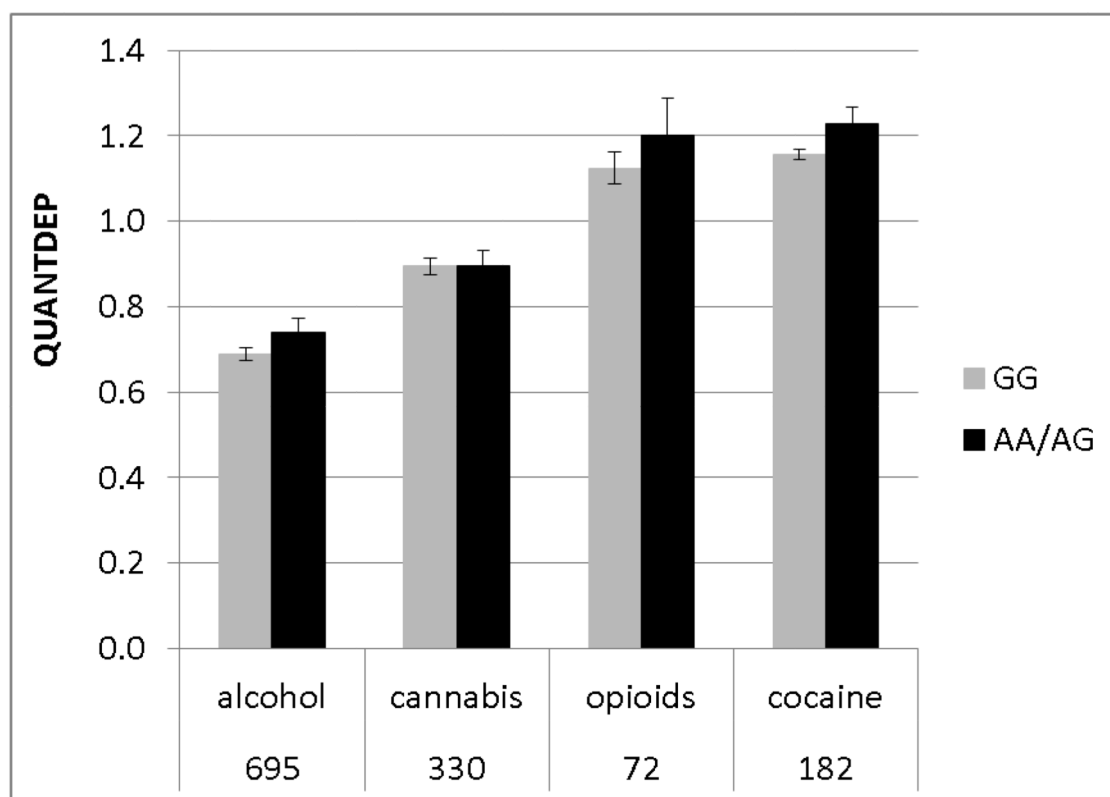




**Figure 2.**

Association results: A. ANYDEP, genomic-control p-values; B. QUANTDEP.

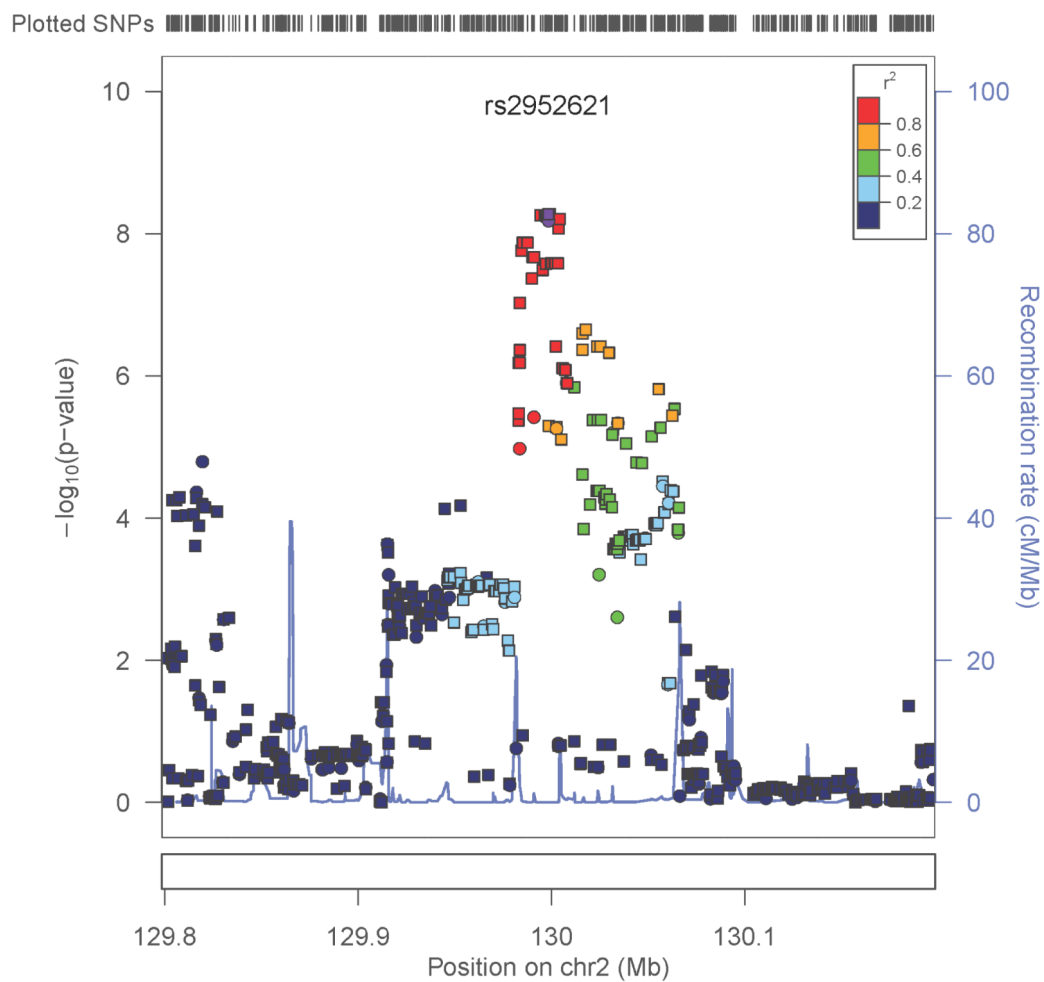


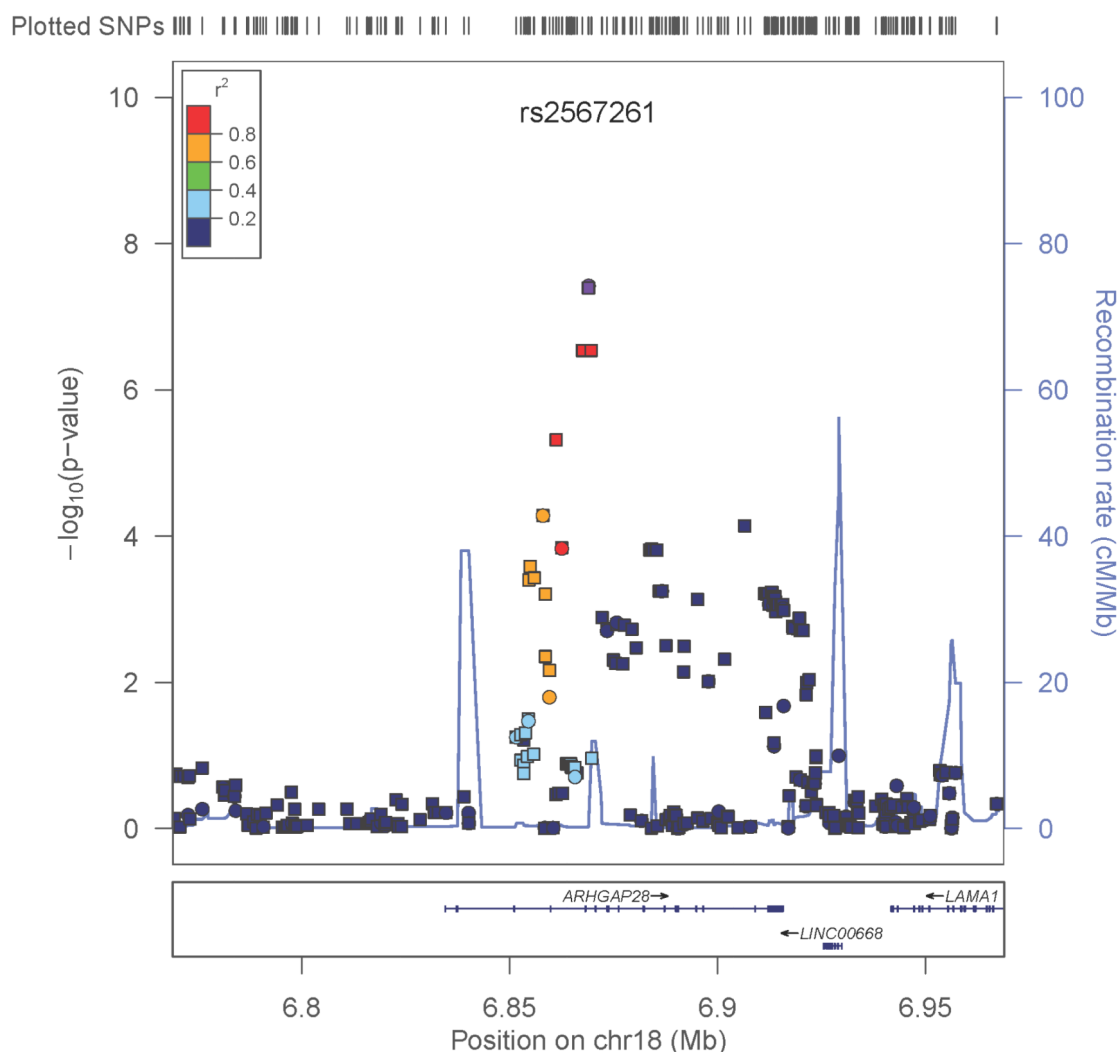


**Figure 3.**

Sample characterization by SNP genotype. A: Dependence on each substance by genotype of rs2952621 in *LOC151121* for all individuals meeting criteria for alcohol, cannabis, cocaine, and opioid dependence; 3B: Mean QUANTDEP by genotype of rs2567261 in *ARHGAP28* for all individuals meeting criteria for alcohol, cannabis, cocaine, and opioid.







**Figure 4. Association results**

A. ANYDEP with genotyped and imputed SNPs in the region flanking *LOC151121*; Y-axis denotes the  $-\log_{10}(\text{p-value})$  for association. X-axis is the physical position on the chromosome (Mb). The most significantly associated SNP is shown in purple. The extent of linkage disequilibrium (as measured by  $r^2$ ) between each SNP and the most significantly associated SNP is indicated by the color scale at top right. Larger values of  $r^2$  indicate greater linkage disequilibrium. Genotyped SNPs are indicated as circles, and imputed SNPs by squares. B. QUANTDEP with genotyped and imputed SNPs in the region flanking *ARHGAP28*. Y-axis denotes the  $-\log_{10}(\text{p-value})$  for association. X-axis is the physical position on the chromosome (Mb). The most significantly associated SNP is shown in purple. The extent of linkage disequilibrium (as measured by  $r^2$ ) between each SNP and the most significantly associated SNP is indicated by the color scale at top right. Larger values of  $r^2$  indicate greater linkage disequilibrium. Genotyped SNPs are indicated as circles, and imputed SNPs by squares.

**Table 1**  
**Substance Dependence for the 832 genotyped individuals meeting any criteria for DSM-IV substance dependence**

SUBSTANCE	# Affected	% Affected of all affected
ANY Alcohol *	696	83.6
ANY Cannabis *	331	39.7
ANY Cocaine *	182	22.4
ANY Opioids *	72	8.9
<b>PATTERNS OF SUBSTANCE DEPENDENCE COMORBIDITY</b>		
Alcohol Alone	405	48.7
Cannabis Alone	85	10.2
Cocaine Alone	23	2.8
Opioids Alone	7	0.8
Alcohol + Cannabis	126	15.1
Alcohol + Cocaine	40	4.8
Alcohol + Opioids	15	1.8
Cannabis + Cocaine	14	1.7
Cannabis + Opioids	2	0.2
Cocaine + Opioids	1	0.1
Alcohol + Cannabis + Cocaine	67	8.0
Alcohol + Cannabis + Opioids	10	1.2
Alcohol + Cocaine + Opioids	10	1.2
Cannabis + Cocaine + Opioids	4	0.5
Alcohol + Cannabis + Cocaine + Opioids	23	2.8
<b>Total</b>	<b>832</b>	<b>100%</b>

\* includes comorbid individuals

**Table 2**  
**Summary of association results for SNPs with  $p < 10^{-6}$  for any trait. All SAGE results are reported. Genome-wide significant results are in bold**

SNP <sup>b</sup>	rs2952621	rs1506807	rs2567261	rs6519647
Chr <sup>a</sup>	2	4	18	22
Position <sup>c</sup>	129,998,443	178,317,692	6,868,925	26,713,964
MAF <sup>d</sup>	0.47	0.25	0.08	0.44
Gene <sup>e</sup>	<i>LOC151121</i> (1kb)	<i>NEIL3</i> (33kb) <i>AGA</i> (34kb)	<b>ARHGAP28</b>	<i>SEZ6L</i>
COGA family p-value for ANYDEP	<b>1.77E-08</b>	9.12E-04	3.18E-04	4.62E-07
SAGE p-value for ANYDEP	0.02	0.63	0.08	0.81
COGA family p-value for QUANTDEP	3.85E-05	4.43E-07	<b>3.76E-08</b>	2.71E-04
SAGE p-value for QUANTDEP	0.06	0.90	0.29	0.73

<sup>a</sup>Chromosome;  
<sup>b</sup>Single Nucleotide Polymorphism  
<sup>c</sup>Chromosomal position (base pairs) based on human genome build 19, dbSNP 137  
<sup>d</sup>Minor allele frequency estimated on founders;  
<sup>e</sup>Gene name and (distance to nearest gene)